



Contribution ID: 74

Type: **Presentation**

On-demand cloud-based secure environments for analysing personal and health data

Wednesday 8 March 2023 11:15 (15 minutes)

Galaxy is the de facto standard workflow manager for bioinformatics providing a complete collaborative platform for researchers. Even though several Galaxy public servers are currently available, there are some situations where users would benefit more from having full administrative control over a private Galaxy instance. These situations include, but are not limited to, worries about data privacy, the need for customization, the need to prioritise particular job types, the development of tools, and training activities.

The Laniakea [1] software platform facilitates the provisioning of on-demand Galaxy instances over heterogeneous Cloud infrastructures, by leveraging on the open source INDIGO-DataCloud cloud stack [2], which aims to make cloud infrastructures more accessible by scientific communities.

End users interact with Laniakea through a web front-end that allows a general setup of the Galaxy instance. The deployment of the virtual hardware and of the Galaxy software ecosystem is subsequently performed by the INDIGO Platform as a Service layer. At the end of the process, the user gains access to a private, production-grade, fully customizable, Galaxy virtual instance. Laniakea features the deployment of stand-alone or cluster backed Galaxy instances, shared reference data volumes, and rapid development of novel Galaxy flavours for specific tasks.

Moreover, to extend the usage of this platform in clinical scenarios, where the analysis of sensitive data, in compliance with the GDPR, requires strong countermeasures to grant data privacy and security, Laniakea guarantees the creation of isolated and secure environments, exploiting storage encryption and access control to Galaxy through VPN, in order to carry out data analysis.

Laniakea allows the on-demand encryption of the entire storage volume attached to the virtual machine, using the Linux kernel encryption module. The level of disk encryption is completely transparent to software applications, in this case Galaxy: data are encrypted and decrypted on-the-fly when writing and reading, respectively. The procedure has been completely automated through the web Dashboard of the PaaS orchestration service [3], taking advantage of Hashicorp Vault for storing user passphrases.

We have implemented a robust mechanism to create secure encryption keys and prevent user credentials or the encryption passphrase from being transmitted unencrypted to the virtual infrastructure, compromising its security.

The oral contribution will provide details about the platform architecture and the service implementation strategy.

References

- [1] Tangaro et al., Laniakea: an open solution to provide Galaxy “on-demand” instances over heterogeneous cloud infrastructures, *GigaScience*, Volume 9, Issue 4, April 2020, g10033, <https://doi.org/10.1093/gigascience/giaa033>
- [2] Salomoni, D., Campos, I., Gaido, L. et al. INDIGO-DataCloud: a Platform to Facilitate Seamless Access to E-Infrastructures. *J Grid Computing* 16, 381–408 (2018). <https://doi.org/10.1007/s10723-018-9453-3>
- [3] <https://github.com/indigo-dc/orchestrator>

Authors: Prof. ZAMBELLI, Federico (Department of Biosciences, University of Milan); DONVITO, Giacinto (INFN - National Institute for Nuclear Physics); TANGARO, Marco Antonio (Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies); ANTONACCI, Marica (INFN); FOGGETTI, Nadina

Presenter: TANGARO, Marco Antonio (Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies)

Session Classification: Security and Authentication

Track Classification: Main session: Technology & Research